

# Περιεχόμενα

Κατάλογος Σχημάτων	xii
Κατάλογος Πινάκων	xviii
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Κίνητρα για Αποθήκες και Εξόρυξη Δεδομένων	2
1.2 Αποθήκες Δεδομένων και Εξόρυξη Δεδομένων	5
1.3 Διαδικασία KDD	7
1.4 Λειτουργίες Εξόρυξης και Εφαρμογές	9
1.5 Σκοποί του Βιβλίου	10
1.6 Ιστορικό και Βιβλιογραφία	11
<b>2 Επεξεργασία Δεδομένων</b>	<b>13</b>
2.1 Μορφή και Ιδιότητες Δεδομένων	14
2.2 Περίληψη Δεδομένων	15
2.3 Καθαρισμός και Μετασχηματισμός Δεδομένων	18
2.3.1 Καθαρισμός δεδομένων	19
2.3.2 Μετασχηματισμός δεδομένων	25
2.4 Κβαντοποίηση Αριθμητικών Ιδιοτήτων	27
2.4.1 Αλγόριθμοι κβαντοποίησης	27
2.5 Τυχαία Δειγματοληψία	32
2.5.1 Απλή τυχαία δειγματοληψία	33
2.5.2 Στρωματοποιημένη δειγματοληψία	38
2.6 Μείωση Αριθμού Διαστάσεων	40
2.6.1 Ανάλυση Πρωταρχικών Συνιστωσών	41
2.6.2 Επιλογή χαρακτηριστικών	46
2.7 Μέτρα Ομοιότητας και Απόστασης	49
2.7.1 Ιδιότητες μέτρων ομοιότητας και απόστασης	49

2.7.2	Γνωστά μέτρα ομοιότητας και απόστασης . . . . .	51
2.8	Ασκήσεις . . . . .	59
<b>3</b>	<b>Αποθήκες Δεδομένων και OLAP</b>	<b>61</b>
3.1	Κίνητρα Ανάπτυξης Αποθηκών Δεδομένων . . . . .	62
3.2	Ορισμός Αποθήκης Δεδομένων . . . . .	65
3.3	Αρχιτεκτονική Αποθήκης Δεδομένων . . . . .	67
3.3.1	Επιχειρησιακές βάσεις δεδομένων . . . . .	67
3.3.2	Περιοχή προετοιμασίας δεδομένων . . . . .	67
3.3.3	Περιοχή παρουσίασης δεδομένων . . . . .	70
3.3.4	Εργαλεία προσπέλασης δεδομένων . . . . .	71
3.4	Κύβος Δεδομένων και Σχήμα Αστέρα . . . . .	71
3.4.1	Κύβος δεδομένων . . . . .	71
3.4.2	Σχήμα αστέρα . . . . .	73
3.5	Πράξεις OLAP . . . . .	74
3.6	Ασκήσεις . . . . .	77
<b>4</b>	<b>Μοντελοποίηση με Διαστάσεις</b>	<b>79</b>
4.1	Τέσσερα Σχεδιαστικά Βήματα . . . . .	80
4.1.1	Σχήμα Αστέρα . . . . .	83
4.1.2	Σχήματα Χιονονιφάδας και Γαλαξία . . . . .	84
4.2	Ιδιότητες Πινάκων Διαστάσεων . . . . .	84
4.2.1	Ονόματα ιδιοτήτων και επιλογή κλειδιών . . . . .	84
4.2.2	Ιεραρχίες και αποκανονικοποίηση . . . . .	88
4.3	Τρεις Τύποι Πινάκων Γεγονότων . . . . .	89
4.3.1	Πίνακες γεγονότων συναλλαγών . . . . .	90
4.3.2	Πίνακες γεγονότων περιοδικής φωτογράφισης . . . . .	91
4.3.3	Πίνακες γεγονότων συσσωρευτικής φωτογράφισης . . . . .	93
4.3.4	Σύγκριση τριών τύπων πινάκων γεγονότων . . . . .	95
4.4	Σχεδιασμός Σύμφωνων Πινάκων Διαστάσεων . . . . .	96
4.5	Μεταβολές στους Πίνακες Διαστάσεων . . . . .	98
4.5.1	Αργά μεταβαλλόμενοι πίνακες διαστάσεων . . . . .	98
4.5.2	Γρήγορα μεταβαλλόμενοι πίνακες διαστάσεων . . . . .	101
4.6	Ασκήσεις . . . . .	102
<b>5</b>	<b>Φυσικό Επίπεδο Αποθηκών Δεδομένων</b>	<b>105</b>
5.1	Τύποι Συστημάτων OLAP . . . . .	106
5.1.1	Συστήματα MOLAP . . . . .	106
5.1.2	Συστήματα ROLAP . . . . .	107

5.1.3	Συστήματα HOLAP . . . . .	108
5.2	Υλοποίηση Κυβοειδών σε Κύβους Δεδομένων . . . . .	108
5.2.1	Κυβοειδή και ο τελεστής CUBE . . . . .	108
5.2.2	Υλοποιημένα κυβοειδή . . . . .	110
5.2.3	Επιλογή κυβοειδών προς υλοποίηση . . . . .	113
5.3	Δομές Καταλόγου για OLAP . . . . .	116
5.3.1	Κατάλογοι bitmap . . . . .	117
5.3.2	Κατάλογοι σύνδεσης . . . . .	118
5.4	Ασκήσεις . . . . .	121
<b>6</b>	<b>Κατηγοριοποίηση</b>	<b>123</b>
6.1	Μοντέλο Κατηγοριοποίησης . . . . .	124
6.2	Τύποι Κατηγοριοποιητών . . . . .	125
6.3	Κριτήρια Αξιολόγησης Κατηγοριοποιητών . . . . .	129
6.3.1	Ακρίβεια πρόβλεψης μοντέλου . . . . .	129
6.3.2	Ευκολία κατανόησης μοντέλου . . . . .	131
6.3.3	Κλιμάκωση μεγέθους συνόλου εκπαίδευσης . . . . .	131
6.3.4	Ανοχή σε θόρυβο και ελλιπείς τιμές . . . . .	132
6.4	Βασικοί Κατηγοριοποιητές . . . . .	132
6.4.1	Δένδρα Απόφασης . . . . .	132
6.5	Bayesian Κατηγοριοποιητές . . . . .	146
6.5.1	Ορισμός Bayesian Κατηγοριοποιητών . . . . .	146
6.5.2	Αφελείς Bayesian Κατηγοριοποιητές . . . . .	148
6.5.3	Bayesian Belief Δίκτυα . . . . .	151
6.5.4	Κατηγοριοποιητές k Πλησιέστερων Γειτόνων . . . . .	152
6.6	Αποτίμηση Ακρίβειας Κατηγοριοποιητών . . . . .	155
6.6.1	Μέθοδοι αποτίμησης ακρίβειας . . . . .	155
6.6.2	Διαστήματα εμπιστοσύνης εκτιμώμενης ακρίβειας . . . . .	157
6.7	Βελτίωση Ακρίβειας Κατηγοριοποιητών . . . . .	158
6.7.1	Υπερβολική προσαρμογή και κλάδεμα . . . . .	158
6.7.2	Σύνολα κατηγοριοποιητών . . . . .	163
6.8	Κλιμακούμενοι Αλγόριθμοι Κατηγοριοποίησης . . . . .	165
6.8.1	Αλγόριθμος SLIQ . . . . .	165
6.8.2	Αλγόριθμος SPRINT . . . . .	170
6.9	Ασκήσεις . . . . .	173

<b>7</b>	<b>Ομαδοποίηση</b>	<b>175</b>
7.1	Ιεραρχική Ομαδοποίηση . . . . .	177
7.1.1	Συναθροιστικός ιεραρχικός αλγόριθμος ομαδοποίησης . . . . .	177
7.1.2	Αλγόριθμοι μονής και πλήρους σύνδεσης . . . . .	180
7.1.3	Γενικευμένα μέτρα απόστασης ομάδων . . . . .	185
7.2	Αλγόριθμοι Τμηματοποίησης . . . . .	189
7.2.1	Απευθείας επίλυση . . . . .	189
7.2.2	Αλγόριθμος k-Means . . . . .	192
7.2.3	Αλγόριθμος Πλησιέστερων Γειτόνων . . . . .	198
7.3	Αλγόριθμος DBScan . . . . .	200
7.4	Αλγόριθμοι Βασισμένοι στη Θεωρία Γράφων . . . . .	204
7.5	Κλιμακούμενοι Αλγόριθμοι Ομαδοποίησης . . . . .	209
7.5.1	Αλγόριθμος BIRCH . . . . .	210
7.5.2	Αλγόριθμος Chameleon . . . . .	212
7.6	Έλεγχος Εγκυρότητας Ομάδων . . . . .	215
7.6.1	Εγκυρότητα με εσωτερικά κριτήρια . . . . .	215
7.6.2	Εγκυρότητα με εξωτερικά κριτήρια . . . . .	222
7.7	Τάση Ομαδοποίησης . . . . .	223
7.8	Ασκήσεις . . . . .	228
<b>8</b>	<b>Συσχέτιση</b>	<b>231</b>
8.1	Κανόνες Συσχέτισης . . . . .	232
8.2	Αλγόριθμος Apriori . . . . .	234
8.2.1	Κριτήριο περικοπής του Apriori . . . . .	235
8.2.2	Αλγόριθμος δημιουργίας και εξέτασης υποψηφίων . . . . .	236
8.2.3	Μέτρηση υποστήριξης . . . . .	239
8.3	Παραγωγή Κανόνων Συσχέτισης . . . . .	241
8.4	Βελτιώσεις Αλγορίθμου Apriori . . . . .	243
8.5	Αλγόριθμος FP-Growth . . . . .	246
8.5.1	FP-δένδρο . . . . .	246
8.5.2	Εύρεση συχνών συνόλων . . . . .	248
8.6	Κλειστά Σύνολα Αντικειμένων . . . . .	251
8.7	Μειονεκτήματα Μέτρου Εμπιστοσύνης . . . . .	253
8.8	Χρήση Περιορισμών . . . . .	256
8.9	Ασκήσεις . . . . .	257

<b>9</b>	<b>Εφαρμογή με Analysis Services</b>	<b>261</b>
9.1	Κύβοι Δεδομένων . . . . .	262
9.1.1	Δημιουργία κύβου . . . . .	262
9.1.2	Προβολή αποτελέσματος . . . . .	292
9.2	Κατηγοριοποίηση . . . . .	293
9.2.1	Δένδρα απόφασης . . . . .	293
9.2.2	Αφελείς Bayesian κατηγοριοποιητές . . . . .	310
9.3	Ομαδοποίηση . . . . .	318
9.3.1	Βασικές ρυθμίσεις . . . . .	318
9.3.2	Παράμετροι . . . . .	322
9.3.3	Προβολή των ομάδων . . . . .	324
9.3.4	Αξιολόγηση του μοντέλου . . . . .	330
9.4	Συσχέτιση . . . . .	334
9.4.1	Βασικές ρυθμίσεις . . . . .	334
9.4.2	Παράμετροι . . . . .	337
9.4.3	Προβολή itemsets . . . . .	339
9.4.4	Προβολή κανόνων . . . . .	342
9.5	Ασκήσεις . . . . .	345
	<b>Βιβλιογραφία</b>	<b>346</b>
	<b>Ευρετήριο</b>	<b>354</b>